

# 40-godišnjica institucije *Cambridge Crystallographic Data Centre* namijenjene pohranjivanju podataka o molekularnim i kristalnim strukturama – “Cambridge Structural Database”

KUI – 14/2006  
Prispjelo 28. studenoga 2005.  
Prihvaćeno 11. travnja 2006.

B. Kojić-Prodić i K. Molčanov

Institut “Ruđer Bošković”, Zavod za fizičku kemiju,  
pp 180, HR-10002 Zagreb, Hrvatska

U 2005. obilježila se 40-godišnjica postojanja i rada svjetski poznatog centra *Cambridge Crystallographic Data Centre* (CCDC), čija je temeljna svrha prikupljanje, provjera i čuvanje kristalografskih podataka koji uz bibliografske podatke sadrže i koordinate atoma i iona koji određuju trodimenzionalnu strukturu molekula u kristalu. *Cambridge Structural Database* (CSD), jedna od prvih elektroničkih baza podataka, najveća je i najznačajnija baza kristalografskih podataka na svijetu s gotovo 400.000 pohranjenih struktura. Upotrebu opsežne baze podataka, koja se stalno povećava, nužno prati i razvoj učinkovitih računalnih programa za pretraživanje, analizu i obradu te vizualizaciju. Ključna uloga CSD-a u istraživanjima vezanim uz kristalografiju, kemiju, znanosti o materijalima, fiziku i kemiju čvrstog stanja, znanosti o životu, farmakologiju, posebice u dizajnu lijekova, dokumentirana je u opsežnoj znanstvenoj literaturi. Izuzetna pažnja posvećuje se točnosti pohranjenih podataka i razvoju programske podrške koja omogućuje široku lepezu primjene. Složen i zahtjevan projekt može ostvarivati samo kompetentan tim stručnjaka, a neki zanimljivi podaci ukazuju na strogo znanstven pristup i tradiciju. Namjera ovog prikaza nije samo obilježavanje jubileja već se želi ukazati na velik broj mogućnosti korištenja baze u istraživanjima i nastavi. Upotreba elektroničkih medija i računalne grafike čini “data mining” ne samo učinkovitim i korisnim, već vizuelno estetskim doživljajem, koji nam pruža arhitektura molekula.

Na Institutu “Ruđer Bošković” već 20 godina uspješno djeluje nekad jugoslavenski, a sada hrvatski nacionalni centar za suradnju sa CCDC-om, koji je i danas odgovoran za akademske korisnike u Hrvatskoj, Sloveniji i Makedoniji omogućavajući im pristup bazi i pratećim računalnim programima.

Upotreba baze u istraživanju prikazana je i vlastitim rezultatima na studiju esterskog kisika kao akceptora protona u nastajanju vodikove veze te topološke analize tubularnih ansambala makrocikličkih politiana.

Ključne riječi: *Kristalografske baze, CCDC, CSD, kristalna struktura, molekularna struktura, molekularno modeliranje*

## Prošlost, sadašnjost i budućnost

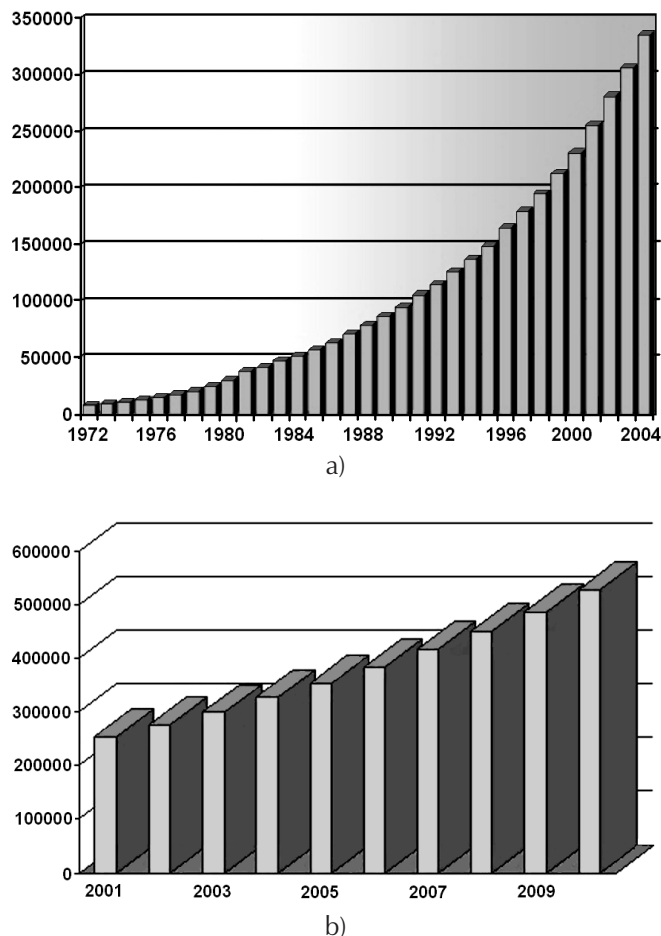
Jednostavan i brz pristup aktualnim i točnim informacijama presudan je za učinkovito odvijanje svih ljudskih djelatnosti. U znanosti su informacije i činjenice temelj na kojem počiva svako istraživanje. Knjižnice se brinu o tiskanim časopisima, knjigama i enciklopedijama te korištenju elektroničkih časopisa i baza podataka putem interneta. U tijeku su opsežne akcije da se putem interneta omogući praćenje što većeg broja časopisa. Obilje informacija dostupno je širokom krugu zainteresiranih čitalaca, kojima je veoma važno da koriste provjerene i točne informacije. U ovom prikazu govorimo o bazi podataka koja je organizirana na način da pruža provjerene i pouzdane podatke.

Prof. dr. Olga Kennard, FRS (članica Kraljevske akademije) utemeljila je 1965. *Cambridge Crystallographic Data Centre*

(CCDC), kao neovisnu jedinicu pri Kemijskom odsjeku Sveučilišta u Cambridgeu, kojoj je temeljni zadatak bio sastaviti bazu kemijskih i kristalografskih podataka za organske spojeve i metalne komplekse s organskim ligandima izučavanjem eksperimentalnim metodama: rendgenskom i neutronskom difrakcijom. CSD ne pohranjuje podatke o strukturama polipeptida i polisaharida s više od 24 podjedinice; njihovi se podaci pohranjuju u *Protein Data Bank*<sup>1</sup> (PDB, <http://www.rcsb.org/pdb>). CSD također ne uključuje strukturne podatke: oligonukleotida, koji se nalaze u *Nucleic Acid Data Bank*<sup>2</sup> (NDB, <http://ndbserver.rutgers.edu>); anorganskih struktura, koje ćemo naći u *Inorganic Crystal Structure Database*<sup>3</sup> (ICSD, <http://www.fiz-informationsdienste.de/en/DB/icsd>); za metale i slitine, čiji su podaci u bazi CRYSMET<sup>4</sup> (<http://www.tothcanada.com>). Na samom je početku djelovanja CCDC-a nastala *Cambridge Structural*

Database<sup>5-7</sup> (CSD, <http://www.ccdc.cam.ac.uk>) koja sadrži koordinate atoma u kristalnoj strukturi kao i osnovne bibliografske, kemijske i kristalografske podatke za sve riješene kristalne strukture. CSD je počeo s podacima za oko 1500 kristalnih struktura s godišnjim povećanjem baze za nekoliko stotina novih struktura. Od skromnih početaka do današnjih dana CSD odražava napredak kristalografije, što se ogleda u golemom povećanju pohranjenih struktura s godišnjim rastom od blizu 30 tisuća (slika 1) te će u ovom desetljeću dostići preko pola milijuna struktura. CSD je neusporedivo najveća i najznačajnija baza eksperimentalno određenih kristalnih struktura na svijetu. Razvoj tehnologije koja unapređuje i usavršava rendgenske difraktometre te razvoj elektroničkih računala, ima presudan utjecaj na usavršavanje eksperimentalnih metoda i tehnika. Posebice su značajne mogućnosti korištenja intenzivnih izvora zračenja – sinkrotrona, korištenje krio-tehnika za mjerenja na temperaturi tekućeg dušika ili helija kao i postojanje osjetljivih i brzih detektora (koji se među ostalim koriste i za svemirske letjelice). Takav napredak u eksperimentalnim tehnikama i metodama rezultira sve većim brojem podataka. Broj riješenih struktura pohranjenih u CSD-u udvostručuje se svakih sedam do osam godina. Većina relevantnih znanstvenih časopisa zahtijeva od autora da kristalografske rezultate pohrane u jednu od kristalografskih baza podataka<sup>7</sup> prije objavljivanja rada; CSD pohranjuje i provjerava kristalografske podatke za 51 svjetski znanstveni časopis, ali izravno i za individualne znanstvenike. Baza pruža aktualan pregled i točne podatke. Elektronički prijenos podataka osigurava golemu propusnu moć i točnost provjere podataka koja je uvedena odmah nakon pojave prve računalne mreže. U tu svrhu razrađen je format za uniformno pohranjivanje podataka i jednostavan elektronički prijenos i kontrolu nazvan CIF (Crystallographic Information File),<sup>8,9</sup> koji se neprestance usavršava. Upravo je u tijeku izrada njegove tehnički savršenije inačice.

Novi ili nedovoljno upućeni korisnici, koji još nisu upoznali sve mogućnosti primjene baze i pripadajuće programske podrške teško mogu predočiti koliko se znanja i rada ulaže godinama u taj projekt. Samo sustavno ulaganje u kreativnost, znanje i trud tima od pedesetak ljudi, tijekom 40 godina, na čelu kojeg je istaknuti znanstvenik prof. dr. Frank H. Allen, objašnjava kvalitetu projekta. Prof. dr. F. H. Allen od 1970. radi u CCDC-u na razvoju računalnih programa i na znanstvenoj aplikaciji baze. Objavio je više od 200 znanstvenih članaka i 16 referentnih knjiga. Za postignuća mu je ukazana čast da 1992. bude izabran za člana "The Royal Society of Chemistry". Također je dobitnik dviju prestižnih nagrada: "The RSC Silver Medal and Prize for Structural Chemistry" 1994. i "Herman Skolnik Award of the American Chemical Society, Division of Chemical Information", 2003. Vrlo opsežan i znanstveno zahtjevan posao obavljao je kao urednik časopisa *Acta Crystallographica, sekcija B* u razdoblju od 1994. – 2002. Uz prof. dr. Allena djeluje pet direktora od kojih svaki ima definirano područje rada: prof. dr. Robin Taylor – direktor razvoja, dr. Steve Salisbury – poslovni direktor, Owen Johnson – tehnički direktor, dr. Sam Motherwell – odgovoran za istraživanje i John Liebeschuetz – odgovoran za primjenu i tržište. Svaki od direktora ima impresivan kurikulum, počevši od diploma i doktorata na prestižnim sveučilištima pa do odgovornih funkcija i priznanja. Spomenut ćemo samo neke podatke koji jasno ukazuju na njihovo široko obrazovanje i stečena znanja. Prof. R. Taylor diplomirao je na Sveučilištu u Oxfordu i doktori-



Slika 1 – a) Porast broja struktura pohranjenih u CSD-u od 1970. do danas, b) očekivani porast broja struktura u CSD-u u sljedećih pet godina

Fig. 1 – a) Growth of number of structures within CSD since 1970, b) expected growth of number of structures in the CSD in the next five years

rao na Sveučilištu u Cambridgeu. Prije no što se pridružio timu CCDC-a proveo je gotovo deset godina na modeliranju za industrijska istraživanja u *Zeneca Agrochemicals*. Dr. Sam Motherwell, starosjedilac u timu, djeluje od 1968. i odgovoran je za korištenje CSD-a u istraživanjima koja se odvijaju jednako intenzivno u okviru samog centra, ali isto tako s neovisnim istraživačima. U posljednje vrijeme intenzivno se zanima za korištenje baze u kristalnom inženjerstvu i predviđanju kristalnih struktura. Od 2002. uključen je u suradnju *Pfizer Institute for Pharmaceutical Materials Science*, CCDC-a i Sveučilišta u Cambridgeu. Svojedobno je djelovao kao voditelj automatizacije u Sveučilišnoj biblioteci u Cambridgeu. Možda će se neke učiniti da su ovo sve suvišni detalji. Radeći s bazom i komunicirajući s ljudima iz Cambridgea, shvatili smo kako je upravo to izvrstan primjer koji potvrđuje izreku da iza svakog dobrog i uspješnog projekta stoje kompetentni stručnjaci.

## Što nam nudi baza

Arhitektura baze opisana je u prikazu objavljenom u *Kemiji u industriji* 2002.<sup>10</sup> Stoga ćemo u ovom osvrtu posvetiti pažnju primjeni baze u rješavanju znanstvenih problema.

Korištenje baze u bibliografske svrhe svodi se na provjeru kojem spoju je poznata trodimenzionalna struktura i koji su je autori objavili s detaljnim podacima o izvoru. Međutim, široka primjena CSD-a u znanstvene svrhe temelji se na podacima o molekularnoj i kristalnoj strukturi spoja, a bitne vrijednosti su koordinate atoma iz kojih možemo izračunati željene geometrijske parametre koji definiraju prirodu kemijske veze između atoma, konformaciju molekula i apsolutnu konfiguraciju, interakcije između molekula u kristalu kao što su vodikove veze, C–H $\cdots$  $\pi$ ,  $\pi\cdots\pi$  i agostičke interakcije te Coulombove interakcije. Poznate koordinate atoma omogućavaju nam da pripadnim računalnim programima grafički prikazemo trodimenzionalnu strukturu molekula i njihovo slaganje u kristalu u skladu sa simetrijom kristala i simetrijom njegove unutarnje građe. Kako su kemijska, fizička i biološka svojstva spojeva određena njihovom elektronskom i molekularnom strukturom, jasno je da je za praćenje odnosa strukture te svojstava i aktivnosti molekula kao i za dizajn novih molekula *Cambridge Structural Database* nepresušno vrelo informacija. U QSAR-u (*Quantitative Structure Activity Relationship*), SAR-u (*Structure Activity Relationship*) i nizu njihovih modifikacija CSD daje realne 2D i 3D strukture molekula (slika 2). Kristalne strukture spojeva daju podatke o interakcijama ne samo među atomima već i molekulama i danas su koristan "alat" u kristalnom inženjerstvu i supramolekularnoj kemiji u sintezi novih nanomaterijala. U prirodi funkcioniraju takvi nanosustavi i čovjeku preostaje da samo imitira prirodu i prepozna gdje će učinkovito primijeniti svoj surogat. Time rengenska strukturna analiza nije samo točna analitička metoda već i alat u sintezi novih spojeva. Direktnim pristupom kombiniramo fotosintezu u čvrstom stanju pomoću lasera na samom difrakcijskom uređaju. Mjerenjima difrakcijskih intenziteta kratkoživućih molekula koje predstavljaju prelazna stanja spoznajemo mehanizam kemijske reakcije. Time od prepoznavanja i bilježenja statičkih stanja prelazimo na studij dinamike procesa i time rasvjetljavamo mehanizme reakcija i otkrivamo kako i zašto možemo utjecati na iskorištenje pri kemijskoj sintezi.

Primjenu CSD-a u rješavanju znanstvenih problema opisuje više od 1300 znanstvenih publikacija. Navedemo li neka područja u kojima možemo iskoristiti bazu, podsjetit ćemo

se što svatko od nas može doznati o sustavu koji istražuje. To su:

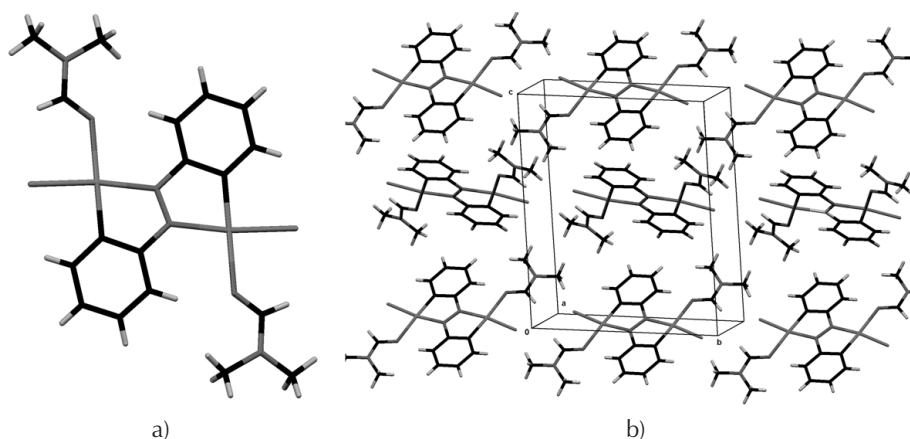
1. Strukturne karakteristike određene klase spojeva;
2. Priroda kemijske veze i elektronski učinci;
3. Kemijska informatika;
4. Kristalno inženjerstvo;
5. Svojstva kristala;
6. Predviđanje kristalne strukture;
7. Simetrija kristala, analiza pojedinih ravnina u kristalu koje su važne za neka fizička svojstva, kao posljedica preferiranih interakcija, predstavljanje teorijskih morfologija kristala: Bravaisove, Friedelove, Donnayeve i Harkerove;<sup>12</sup>
8. Izračunavanja potencijalne energije molekule i energije kristalne rešetke;
9. Vodikova veza i ostale nekovalentne interakcije;
10. Svojstva organskih molekula koje se koriste kao ligandi u pripravi metalnih kompleksa te kao supstrati ili inhibitori enzima, što uključuje izučavanje interakcija između proteina i liganada;
11. Geometrija molekula i stereokemija: konformacijska analiza, sterički i elektronski efekti;
12. Molekularno modeliranje i dizajn lijekova i spojeva zadanih svojstava;
13. Simetrija molekula, informacije za spektroskopska istraživanja;
14. Fizičko-kemijska svojstva molekula;
15. Mehanizmi reakcija i strukturna korelacijska analiza;
16. Statističke metode i analiza pogrešaka.

Kako bismo izdvojili i obradili podatke za istraživanje u navedenim područjima, na raspolaganju je sofisticirana programska podrška koja sistematizira i matematički obrađuje podatke, a na istraživaču je da ih razumije i ispravno interpretira. Navedenih 16 područja pokazuje koliko je univerzalna primjena baze. Prikaz nije usmjeren iscrpnom opisu postojećih računalnih programa, već mu je svrha da odabranim primjerima pokaže njihovo korištenje. Želimo istaknuti da je upoznavanje studenata na dodiplomskom, a svakako na doktorskom studiju, s bazom i njezinom upotrebom korisno iskustvo i svakako ga treba predvidjeti. Autori

članka to već primjenjuju u okviru kolegija Odabrana poglavlja rengenske strukturne analize na Prirodoslovno-matematičkom fakultetu Sveučilišta u Zagrebu.

### Tematski izdvojene baze unutar CSD-a i programi za obradu podataka

Osnovni program za pretraživanje CSD-a od travnja 2000. je *ConQuest*,<sup>13</sup> kojim je zamijenjen stariji *Quest3D*.<sup>14</sup> Zamišljeno je da *ConQuest* jednostavnim grafičkim ulazom postavi pitanje bazi putem dvodimenzionalnog dijagrama molekule. Program je prilagođen računalnim platformama MS WIN-



Slika 2 – a) Prikaz molekulske strukture paladijevog kompleksa  $\{PdCl(dmf)\}_2(\mu-C_6H_4N=NC_6H_4)^{11}$  i b) njegovog kristalnog pakiranja<sup>11</sup> načinjen programom Mercury

Fig. 2 – a) Molecular structure of palladium complex  $\{PdCl(dmf)\}_2(\mu-C_6H_4N=NC_6H_4)^{11}$  and b) its crystal packing prepared<sup>11</sup> by the CCDC program Mercury

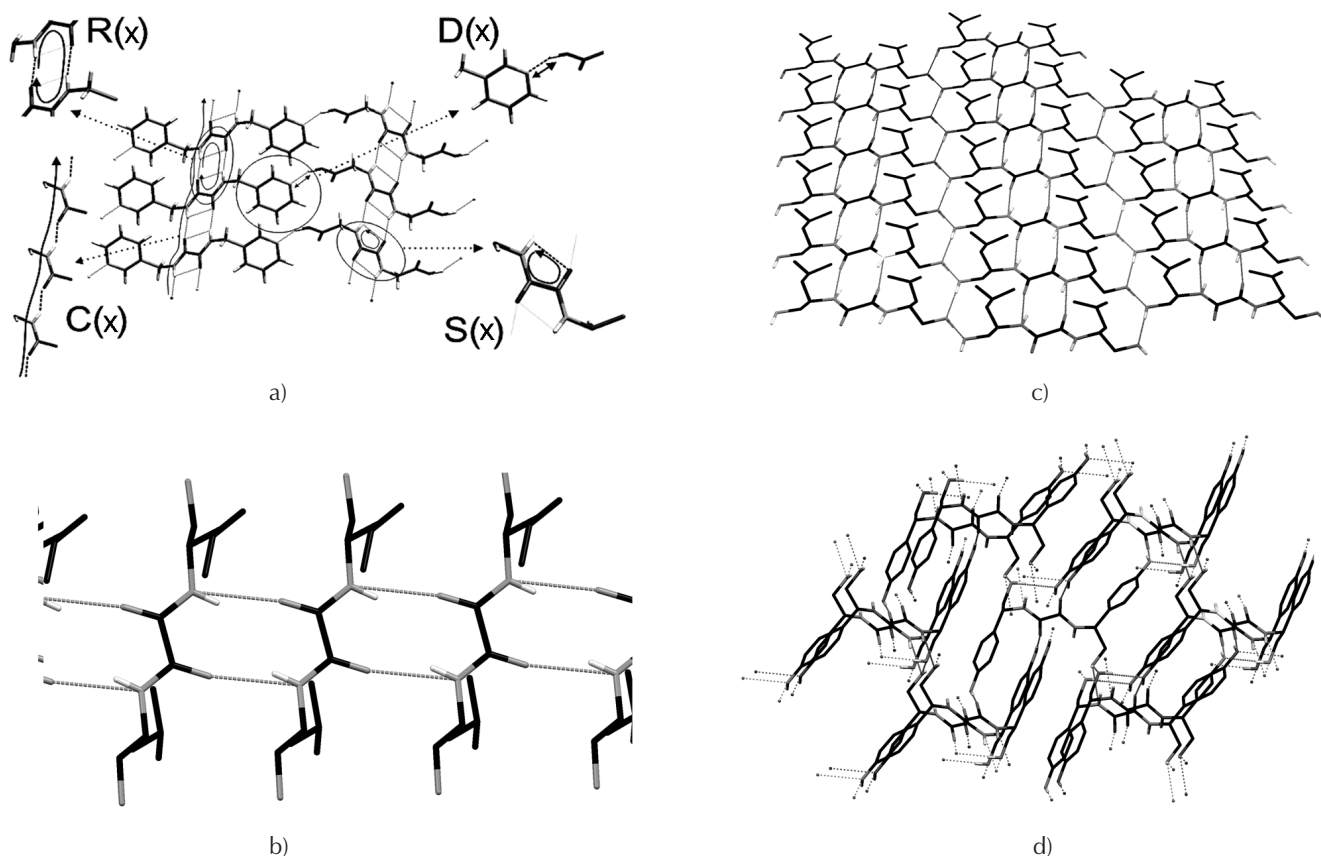


*DOWS* (uz *UNIX* i njegove izvedenice *SGI-IRIX* i *LINUX*). Za statističku analizu brojčanih podataka iz baze primjenjuje se program *Vista* (ili *MS-Excel*). *Vista* se koristi zadanim vrijednostima geometrijskih parametara izabranih kristalnih struktura pohranjenih u *CSD*-u za različite statističke analize. Pomoću tog programa moguće je statistički obraditi odabrane duljine veza, valentnih ili torzijskih kutova i time ispitati utjecaj steričkih i elektronskih efekata ili utjecaj interakcija u kristalu. Sustavnom analizom može se objasniti odstupanja od geometrije predviđene hibridizacijom. Rezultati se mogu predložiti računski i grafički histogramima programom *Vista*.

Najnoviji program za trodimenzionalnu vizualizaciju molekularnih i kristalnih struktura iz *CCDC*-a je *Mercury*, uveden 2001.<sup>13</sup> Taj program omogućava prikaz kristalnog pakovanja i analizu vodikovih veza s oznakama temeljenim na teoriji grafova (slika 3). Također može iz koordinata atoma i pripadnih kristalografskih podataka izračunati rendgeno-

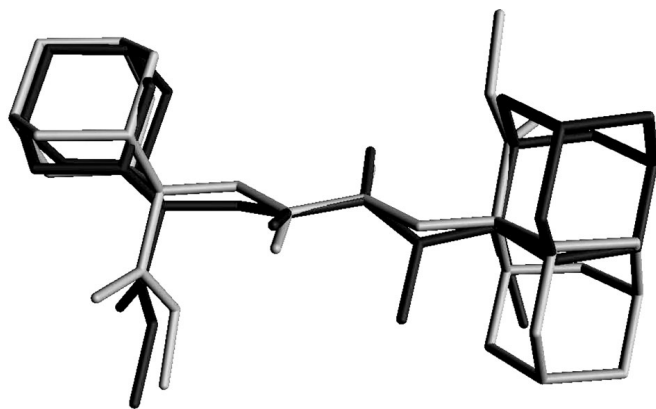
gram praha pojedinog spoja koji pruža veliku pomoć pri analizi sastava smjesa metodom rendgenske difrakcije na praškastom uzorku. Rendgenogram praha služi za identifikaciju kristalnog materijala, pa ta mogućnost baze također ima veliku primjenu. Prilikom analize kristalnog pakiranja može usporediti geometriju kemijski istih podjedinica metodom najmanjih kvadrata i pokazati dijagram prekrivanja (slika 4), što je npr. prikladno ako u strukturi postoji više konformera. Pomoću tog programa možemo pratiti preferirane interakcije u određenim ravninama u kristalu (slika 5), što može biti npr. zanimljivo za kalavost, vodljivost ili neko drugo svojstvo.

Nadgradnju *CSD*-a predstavlja podbaza nekovalentnih interakcija *IsoStar*,<sup>17</sup> koja je uvedena 1997. kao i program *SuperStar*,<sup>18</sup> temeljen na empirijskoj metodi pronalaženja aktivnih mjesta u makromolekulama. *SuperStar*<sup>18</sup> primjenjuje znanje pohranjeno u podbazi nekovalentnih interakcija *IsoStar*<sup>17</sup> utemeljenoj na kemijskim znanjima o međumo-



Slika 3 – a) Osnovni motivi u “graf-set” opisu mreža vodikovih veza: **R** (broj donora, broj akceptora)-motiv prestena, **D** (x)-konačni motiv, **C** (x)-motiv lanca i **S** (x)-intramolekularni motiv koriste se za opisivanje mreže vodikovih veza. Primjer je struktura uzeta iz *CSD* (oznaka svake kristalne strukture u *CSD*-u, tj. pristupni kod “refcode”; pristupni kod ove strukture je CAJQAG). b) Izdvojen jednodimenzionalni motiv vodikovih veza tipa “ljestava” karakterističan za oksalamidne jedinice u kristalnoj strukturi spoja (S,S)-N,N'-oksalil-dileucinola koji uz dodatno povezivanje preko terminalnih skupina tvori u kristalu c) dvodimenzionalnu mrežu vodikovih veza. d) Kristalno pakiranje (S,S)-N,N'-oksalil-dihidroksifenilglicina s trodimenzionalnom mrežom vodikovih veza. Primjeri su preuzeti iz doktorske disertacije dr. Zorana Štefanića, obranjenj u Prirodoslovno-matematičkom fakultetu (smjer Fizika, biofizika) u Zagrebu, prosinca 2004.<sup>15</sup>

Fig. 3 – a) Basic motives in graph-set description of hydrogen-bonded networks: **R** (no. of donors, no. of acceptors)-ring motive, **D** (x)-finite motive, **C** (x)-chain motive and **S** (x)-intramolecular motive are used to describe hydrogen-bonded networks. The example shown is the structure from the *CSD* (*CSD* access code for each set of data, “refcode”; for this structure “refcode” is CAJQAG). b) An isolated “ladder” motif of one-dimensional hydrogen-bonded network, characteristic of oxalamide moieties in the crystal structure of (S,S)-N,N'-oxalyl-dileucinol which forms a c) two-dimensional network by hydrogen bonding of terminal groups. d) Crystal packing of (S,S)-N,N'-oxalyl-dihydroxyphenylglycine with a three-dimensional hydrogen-bonded network. The examples are taken from PhD thesis of dr. Zoran Štefanić at Faculty of Science, University of Zagreb, December 2004.<sup>15</sup>



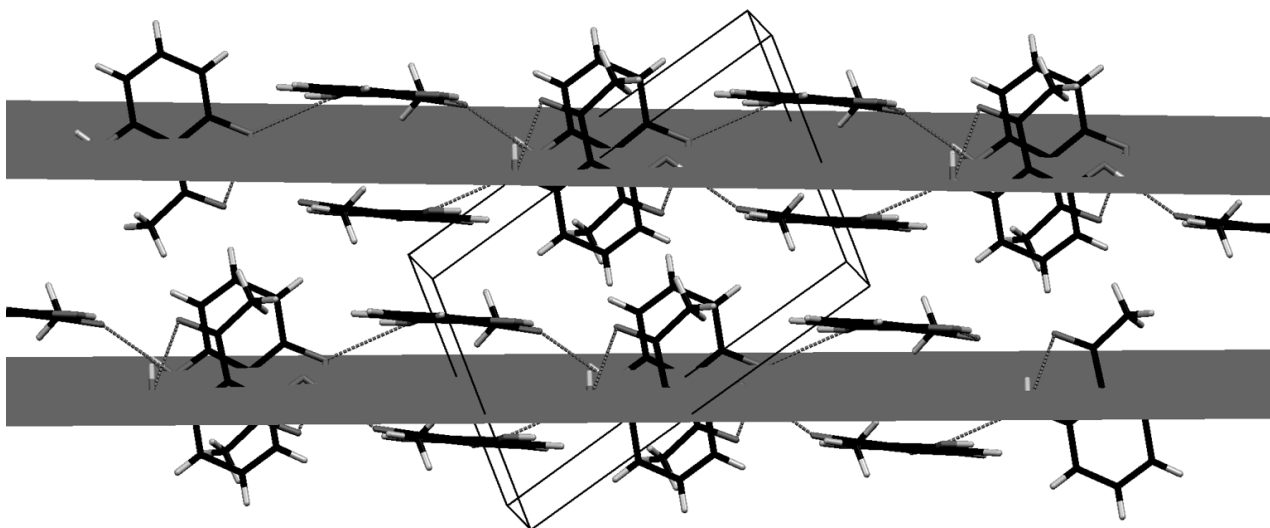
Slika 4 – Prekrivanje dviju kemijski sličnih molekula, od kojih jedna ima molekularnu simetriju  $C_7$  (svjetlija) s anti-rasporedom adamantanskih podjedinica, dok druga ima  $C_1$  (tamnija) sa syn-rasporedom (u pripremi za tisak). Prikaz je načinjen programom Mercury.<sup>13</sup>

Fig. 4 – Overlap of two chemically similar molecules, one of which is  $C_7$ -symmetric (light gray) with anti arrangement of adamantyl moieties, while the other has a  $C_1$  symmetry (dark gray) and syn arrangement of adamantyl moieties (to be published). The figure is prepared with the CCDC program Mercury.<sup>13</sup>

lekularnim, tj. nekovalentnim interakcijama velikog broja funkcionalnih skupina. IsoStar objedinjuje podatke iz CSD-a, PDB-a i rezultate računa molekularnih orbitala temeljene na *ab initio* pristupu primjenom intermolekularne perturbacijske teorije (IMPT).<sup>19</sup> Primjenom te programske podrške možemo analizirati topologiju i geometriju ostvarenu nekovalentnim interakcijama u kristalu. Kao primjer pogledajmo prosječno okruženje molekule kristalne vode (slika 6), tj. prostorni raspored vodikovih veza na temelju pretraživanja velikog broja struktura pohranjenih u CSD-u i PDB-u. Takva analiza pokazuje da molekule (kristalne) vode teže te-

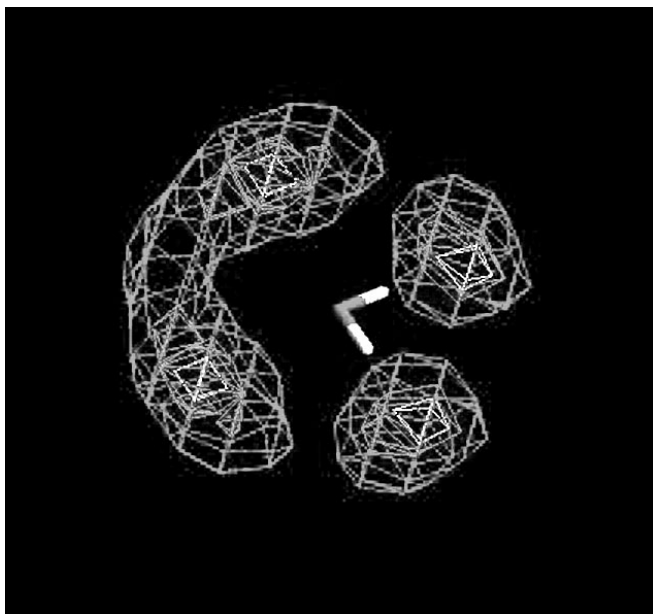
traedarskoj koordinaciji. Svaka nekovalentna interakcija može se detaljno istražiti, uzevši u obzir eksperimentalne podatke o van der Waalsovima radijusima atoma i potom usporediti s teorijskim rezultatima kvantno-mehaničkih proračuna za pojedini modelni sustav. Sustavnim pretraživanjem pomoću IsoStara može se otkriti novi tip interakcija. Podaci pohranjeni u CSD-u mogu koristiti teorijskim kemičarima za provjeru izračunatih modela i studij molekularnih modela i interakcija, račun energije kao npr. energiju optimalne konformacije molekule, elektrostatske energije, energije prijenosa naboja i disperzijske energije. U svakom slučaju, ovaj alat veoma često upotrebljavaju, uz kristalografu, i stručnjaci koji se bave molekularnim modeliranjem.

Među novijim proizvodima CCDC-a je program GOLD,<sup>20,21</sup> koji primjenjuje genetski algoritam razvijen u suradnji sa stručnjacima Glaxo Wellcome i Sveučilišta u Sheffieldu, a korisnicima je stavljen na raspolaganje 1998. godine. GOLD (Genetic Optimization for Ligand Docking)<sup>20,21</sup> je program za predviđanje dinamike i načina vezivanja liganda u aktivno mjesto proteina (*protein docking*) (slika 7). GOLD za razliku od SuperStara ne koristi samo program temeljen na znanju već računa optimalan način vezivanja liganda u aktivno mjesto ciljanog proteina time da koristi polje sila usklađeno s eksperimentalnim podacima iz CSD-a; torzijski kutevi liganada ograničavaju se na područja prepoznata CSD-om, a za usmjerenost nekovalentnih interakcija (vodikove veze) odabiru se one vrijednosti koje su prepoznate u bazi. Način na koji će ligand naći svoj put do aktivnog mjesta od velike je važnosti za istraživanja na polju oblikovanja novih lijekova i proizvodnje učinkovitih i ekološki prihvatljivih agrokemikalija. Zato su stručnjaci CCDC-a i Sveučilišta u Sheffieldu te Glaxo-Welcome razvili algoritam temeljen na pretraživanju konformacijskog prostora uz punu gibljivost liganda i djelomičnu konformacijsku pokretljivost samog proteina. Taj algoritam uvažava konformacijske i fiziološke značajke prirodnog sustava. GOLD je provjeren



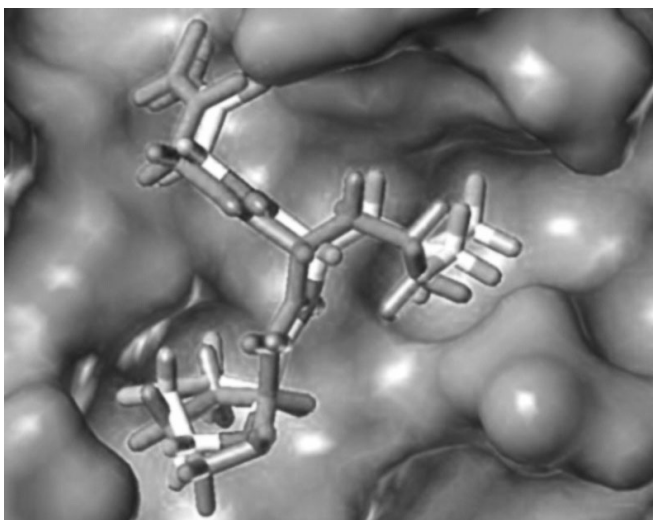
Slika 5 – Dvodimenzionalna mreža vodikovih veza: prikaz ravnina u kristalnoj strukturi dihidroksiacetofenona<sup>16</sup> načinjen programom Mercury.<sup>13</sup> Spoj se primjenjuje u spektrometriji masa kao matrica za vezivanje analizirane supstance. Nije poznato o čemu ovise dobra svojstva matrice. Pretpostavlja se da matrica s hidrofilnim graničnim slojem "pogoduje" interakcijama s hidrofilnim površinama analizirane supstance, konkretno proteinom. Stoga je korisno istaknuti potencijalne ravnine interakcije koristeći ovu mogućnost programa.<sup>13</sup>

Fig. 5 – Two-dimensional hydrogen-bonded layers: planes in the crystal structure of dihydroxyacetophenone<sup>16</sup> prepared by the program Mercury.<sup>13</sup> The compound is used in mass spectrometry as a matrix. The properties which make a good matrix are not known; it is assumed that the matrix with hydrophilic layers is available for interactions with the hydrophilic surface of an analyte, namely protein.



Slika 6 – Vjerojatnost nalaženja donorskih i akceptorskih skupina oko molekule vode, izveden programom IsoStar.<sup>17</sup> Prosječno okruženje molekula kristalne vode je tetraedarsko, a proton-donori su usmjereni prema slobodnim elektronskim parovima kisikovog atoma.

Fig. 6 – Probability distribution of donor and acceptor moieties around a water molecule, prepared by the program IsoStar.<sup>17</sup> Average coordination of a crystal water molecule is tetrahedral, with hydrogen-bond donors directed towards lone electron pairs of the oxygen atom.

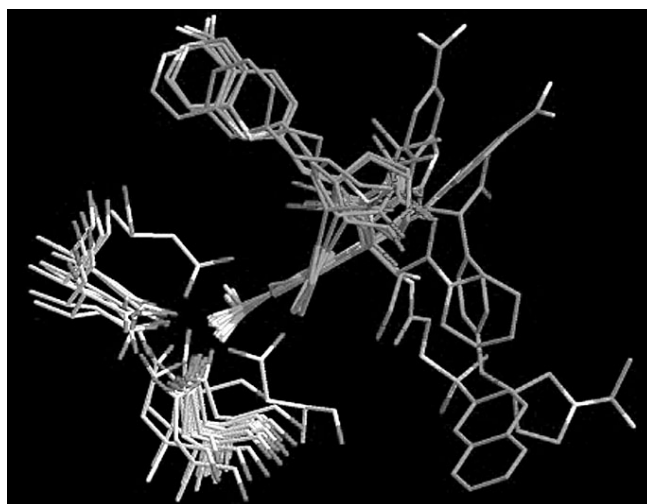


Slika 7 – Usporedba smještanja molekule liganda (modeli od štapića) u aktivno mjesto proteina (pozadina) predviđenog pomoću programa GOLD (molekula bliže promatraču) s eksperimentalno dobivenom kristalnom strukturom (šarena molekula dalje od promatrača). Slika je načinjena pomoću programa GOLD.<sup>20,21</sup>

Fig. 7 – Comparison of docking of a ligand molecule (stick models) in the active place of a protein (background) as predicted by the program GOLD (molecule nearer to the observer) and experimentally obtained crystal structure (coloured by element; farther from the observer). The figure is prepared by the program GOLD.<sup>20,21</sup>

na temelju 136 kompleksa protein-ligand iz *PDB*-a. Analiza je pokazala da se u 71 % slučajeva predviđanje slaže s eksperimentalnim podacima.

Uz to, CCDC u suradnji s dr. Manfredom Hendlichom na Sveučilištu u Marburgu distribuira i razvija *ReLiBase*<sup>22</sup> – za pretraživanje i rukovanje podacima iz *Protein Data Bank* (*PDB*) u studiju interakcija proteina i liganda-male organske molekule koja djeluje kao supstrat ili inhibitor (slika 8). *ReLiBase* se može koristiti na sustavima *WINDOWS* i *UNIX* (uključujući i *LINUX*). Očito je da su ti podaci korisni i nužni u dizajnu novih lijekova jer omogućavaju praćenje potencijalnog lijeka i proteina koji je ključan pri danom oboljenju.



Slika 8 – Rezultat pretrage nekovalentnih interakcija između benzamidinske skupine (liganda) i proteinskih karboksilnih skupina. Benzamidinski fragmenti iz različitih struktura superponirani su da bi se uočila usmjerenost vodikovih veza koje se ostvaruju između karboksilnih skupina male molekule-liganda (lijeka tj. inhibitora ili supstrata) i proteina.

Fig. 8 – Result of a database search of noncovalent interactions between the benzamidine moiety (ligand) and protein carboxyl moieties. Benzamidine fragments from different structures have been superimposed to emphasise directionality of hydrogen bonds between the carboxyl moieties of the small molecule-ligand (drug, i.e. inhibitor or a substrate) and the protein.

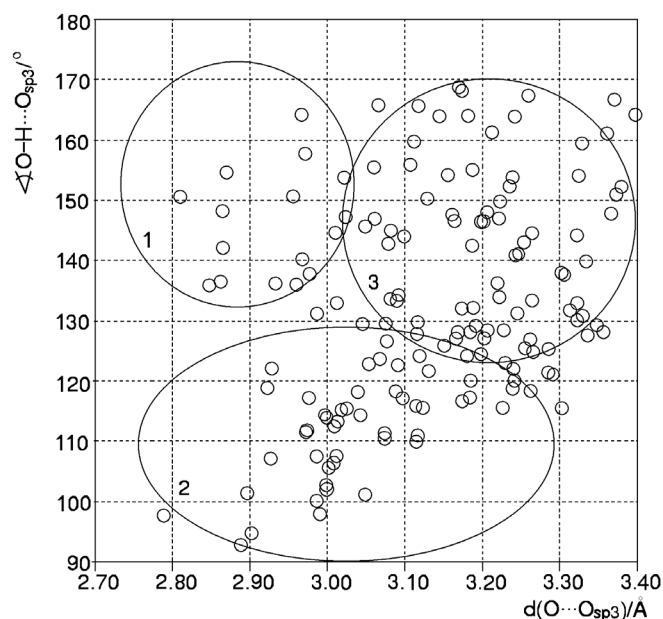
## Primjeri korištenja CSD-a:

### a) Analiza karaktera neuobičajene vodikove veze s esterskim kisikom kao akceptorom

Uloga  $sp^2$  kisika (karbonilna skupina) kao akceptora protona u vodikovoj vezi opsežno je istražena jer je to veoma rasprostranjena interakcija. Naprotiv, esterski  $sp^3$  kisikov atom smatra se slabim akceptorom vodikove veze, i zbog blizine karbonilnog kisikovog atoma posve je nevjerojatno da će sudjelovati u vodikovoj vezi. Lommerse i suradnici su 1997.<sup>23</sup> izučili mogućnosti akceptorske uloge karbonilne, eterske i esterske skupine u vodikovim vezama pristupom *ab initio* i usporedili to s eksperimentalnim podacima iz Cambridge Structural Database. U to vrijeme nisu našli primjere s esterskim  $sp^3$  kao akceptorom. U određivanju kristalnih struktura sreli smo se upravo s takvim primjerima i odlučili analizirati veći broj struktura.<sup>24</sup> Pretraživanjem CSD-a



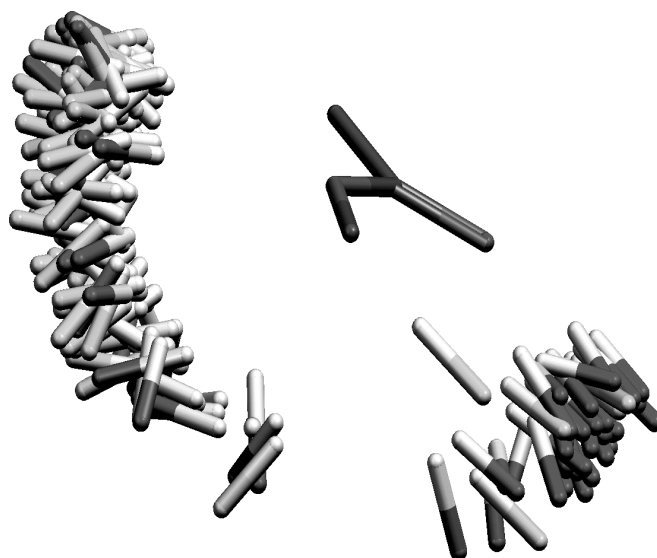
ustanovljeno je da postoji dvjestotinjak kristalnih struktura kod kojih je esterski  $sp^3$  kisikov atom akceptor vodikove veze s kisikovim ili dušikovim atomom kao donatorom. Takve vodikove veze,  $O-H\cdots O_{sp^3}$ , možemo podijeliti na tri grupe:<sup>25,26</sup> srednje jake (grupa 1 na slici 9), račvaste kod kojih je  $O_{sp^3}$  sporedni akceptor (grupa 2 na slici 9) i slabe (grupa 3 na slici 9). Pomoću programa *IsoStar*<sup>17</sup> i *Mercury*<sup>13</sup> mogu se preklopiti odabrani dijelovi svih struktura kod kojih je esterski  $sp^3$  kisikov atom akceptor vodikove veze. Na taj je način načinjena slika 10.



Slika 9 – Prikaz ovisnost duljine vodikove veze (s esterskim  $sp^3$  kisikovim atomom kao akceptorom)<sup>24</sup> o kutu  $O-H\cdots O_{sp^3}$ . Veze kraće od približno 3 Å i s kutovima većim od 130° mogu se smatrati srednje jakim (grupa 1), dok su one s kutovima od 90° – 130° najčešće račvaste (grupa 2). Granicu između slabih veza (grupa 3) i van der Waalsovih interakcija teško je odrediti; one kraće od 3,4 Å obično se smatraju vodikovim vezama. Podaci za grafikon obrađeni su programima *ConQuest*<sup>13</sup> i *Vista*.

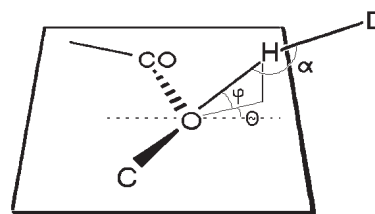
Fig. 9 – Plot of the hydrogen-bond length (with ester  $sp^3$  oxygen atom as an acceptor)<sup>24</sup> vs.  $O-H\cdots O_{sp^3}$  angle. Bonds shorter than ca. 3 Å with angles larger than 130° can be regarded as medium-strong (cluster 1), while those with angles 90° – 130° are mostly bifurcated (cluster 2). The border between weak bonds (cluster 3) and van der Waals interactions is difficult to define; those shorter than 3.4 Å are usually regarded as hydrogen bonds. The data for the graph are processed by programs *ConQuest*<sup>13</sup> and *Vista*.

Usmjerenost prema slobodnim parovima definirat ćemo pseudopolarnim kutom  $\Theta$  kao na slici 11. Ako je vodikov atom usmjeren prema jednom slobodnom elektronskom paru,  $\Theta$  će biti oko 55°, a ako je usmjeren između parova, bit će oko 0°. Kod estera u konformaciji *anti* prevladava usmjerenost prema jednom slobodnom paru, dok je kod konformacije *syn* veza najčešće usmjerena između slobodnih parova. Ti se rezultati slažu s *ab initio* proračunima.<sup>23</sup> Izvršena analiza pokazala je: a) da je taj tip vodikove veze moguć, b) da je uvjetovan steričkim i elektronskim svojstvima skupina koje sudjeluju u vezi i c) razjasnio način kompeticije  $O_{sp^3}$  s polarnijim akceptorima kao što je  $O_{sp^2}$ . Rezultati su detaljno opisani u radu objavljenom u *Acta Crystallographica B*.<sup>24</sup>



Slika 10 – Raspodjela donora protona u vodikovoj vezi oko esterske skupine.<sup>24</sup> Prikaz je pripremljen programom *IsoStar*.<sup>17</sup>

Fig. 10 – Distribution of proton donors around an ester moiety.<sup>22</sup> The figure was prepared by the program *IsoStar*.<sup>17</sup>



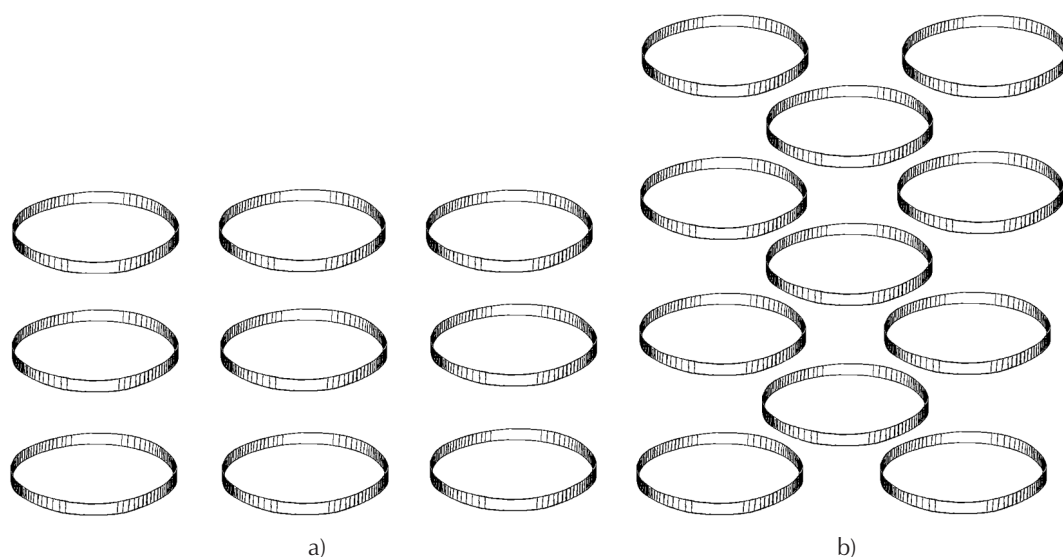
Slika 11 – Definicija pseudopolarnog kuta  $\Theta$ <sup>24</sup>

Fig. 11 – Definition of the pseudopolar angle  $\Theta$ <sup>24</sup>

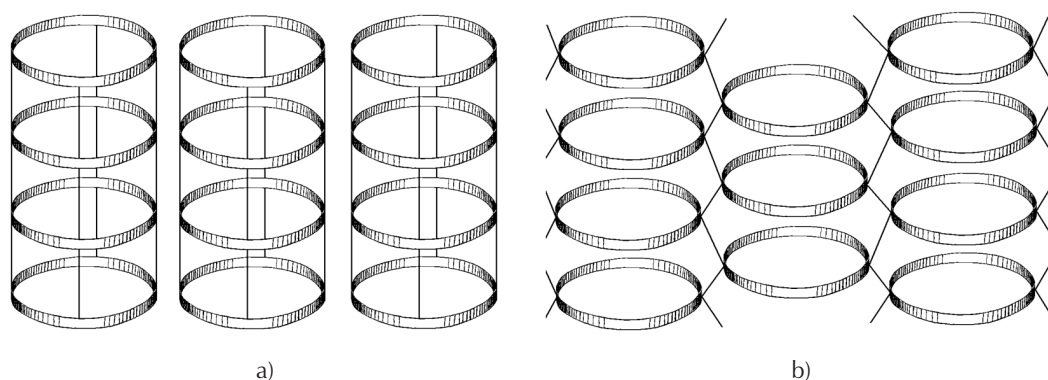
## b) Topološka analiza tubularnih ansambala makrocikla s halkogenim atomima

Kako je stereokemija makrocikla sa sumporom u prstenu samo djelomično opisana, željeli smo izučiti primjere iz baze i sistematizirati stereokemiju i konformaciju te skupine spojeva i objasniti parametre koji određuju nastajanje tubularnih molekulskih ansambala.<sup>27,28</sup> Makrociklički politiani skloni su tvoriti tubularne strukture (slika 12). Od 137 kristalnih struktura cikličkih spojeva s prstenovima većim od 12 atoma kod kojih je prisutan fragment  $C\alpha-S-CH_2-CH_2-S-C\alpha$  60 (tj. više od 40 %) tvori tubularne slagaline. Kod njih se prstenovi najčešće slažu jedan na drugi (52 strukture, slika 12a), a rjeđe (8 struktura) se preklapaju kao na slici 12b. Slabim vodikovim vezama, najčešće tipa  $C-H\cdots S$  i  $C-H\cdots O$ , povezuju se na dva topološki različita načina (slika 13).

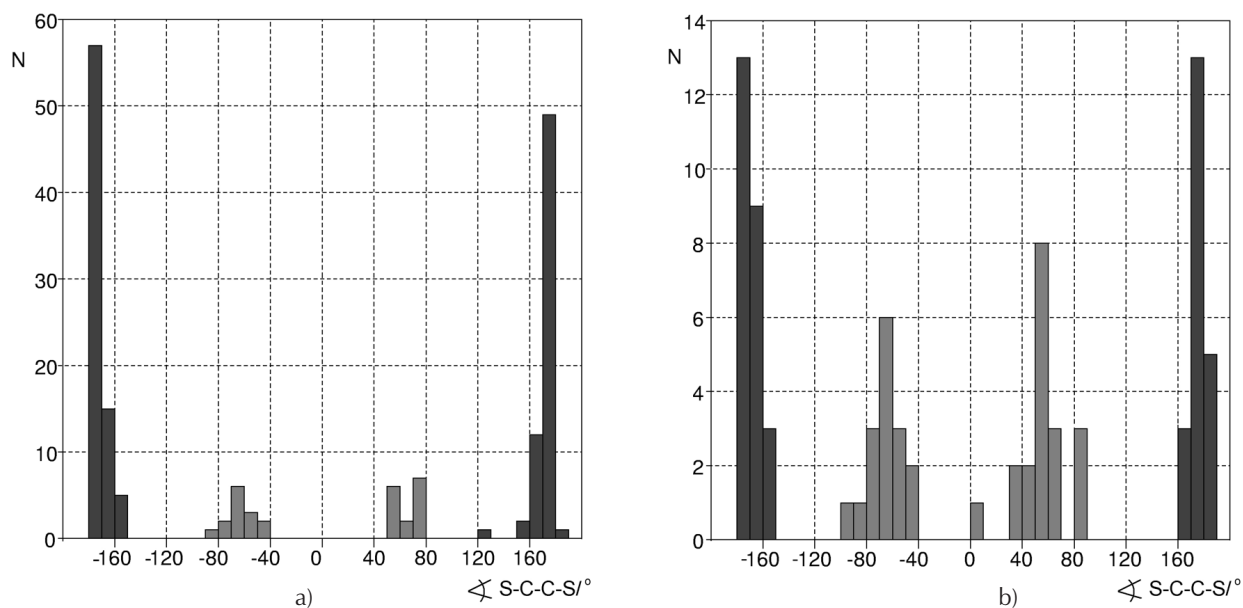
Konformacija skupine  $S-C-C-S$  znatno ovisi o napetosti prstena, ali i o steričkim smetnjama koje stvaraju skupine  $C\alpha$ . Kod velikih prstenova (12 ili više članova) koji nisu sterički ometeni ( $C\alpha$  su metilenske skupine), konformacije su gotovo isključivo *anti* (sl. 14a). U strukturama kod kojih su oba  $C\alpha$  atoma  $sp^2$ -hibridizirani napetost je znatno veća pa se češće javljaju konformacije *gauche* (slika 14b). Kod manjih prstenova (11 ili manje članova) s oba  $C\alpha$  atoma  $sp^2$ -hibridizirana, skupine  $S-C-C-S$  su gotovo sve *gauche*.



Slika 12 – Slaganje prstenova u tubularnim strukturama: a) odvojeni tubuli, b) tubuli s djelomično preklopljenim prstenovima<sup>28</sup>  
 Fig. 12 – Stacking of rings in tubular structures: a) discrete tubules, b) tubules with partially overlapping rings<sup>28</sup>

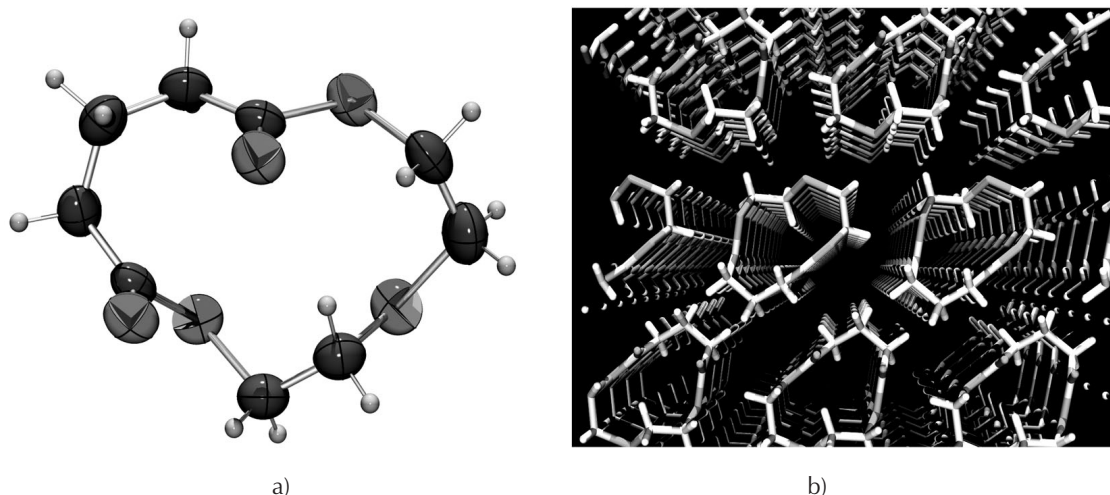


Slika 13 – Vodikove veze između prstenova u tubularnim strukturama: a) povezivanje unutar tubula; b) povezivanje između tubula (cik-cak)<sup>28</sup>



Slika 14 – Raspodjela torzijskih kutova S–C–C–S u kristalnim strukturama makrocikličkih politijana: a) za fragmente  $\text{CH}_2\text{--S--CH}_2\text{--CH}_2\text{--S--CH}_2$ ; b) za fragmente  $\text{Csp}^2\text{--S--CH}_2\text{--CH}_2\text{--S--Csp}^2$ .<sup>28</sup>  
 Fig. 14 – Distribution of torsion angles S–C–C–S in crystal structures of macrocyclic polythianes: a) for fragments  $\text{CH}_2\text{--S--CH}_2\text{--CH}_2\text{--S--CH}_2$ ; b) for fragments  $\text{Csp}^2\text{--S--CH}_2\text{--CH}_2\text{--S--Csp}^2$ .<sup>28</sup>





Slika 15 – a) Molekularna struktura spoja koji tvori tubule;<sup>27,28</sup> b) Slaganje prstenova optimalne konformacije koja dopušta prostorno usmjeravanje podesno za interakciju C–H...O odgovornu za formiranje tubula<sup>27,28</sup>

Fig. 15 – a) Molecular structure of a tubule-forming compound;<sup>27,28</sup> b) Rings of appropriate conformations stack in order to maximise number of C–H...O interactions<sup>27,28</sup>

Analiza pokazuje kako treba usmjeriti sintezu da se dobiju makrocikli optimalnih konformacija, bez steričkih napetosti, koje usmjeravaju donorske i akceptorske skupine u položaj za vodikovu vezu kojom se vjerno kopiraju molekule tvoreći tubularne ansamble (slika 15).<sup>27,28</sup>

## Zaključak

Na temelju pruženih informacija, priloženih literaturnih izvora te primjera iz vlastitog istraživanja čitalac je spoznao kako može koristiti Cambridge Structural Database. Prihvaćamo izazov da navedemo naša očekivanja o ulozi baza podataka sa strukturama molekula u istraživanjima koja će zacijelo mijenjati naše spoznaje. CSD se koristi u predviđanju kristalnih struktura i polimorfije. Međutim to područje još nije postiglo svoj maksimum. Uz sve što smo

ukratko rekli o korištenju baza podataka u znanostima o životu otvara se jedan novi aspekt: povezivanje strukturne sistematike proteina prema njihovoj trodimenzionalnoj građi i funkciji, te bioinformatike i evolucije. Mapiranjem genoma većeg broja organizama, uključujući humani genom, predstoji velik posao u otkrivanju funkcije proteina. U tom smjeru postaje posebno zanimljivo predviđanje svijanja proteina (*folding*) koje je izravno odgovorno za funkciju svakog proteina. Time dobiva još više na značenju sistematika vezana uz strukturne baze. Izuzetno privlačno područje u začetku je znanost informacijskih molekularnih sustava<sup>29</sup> kojoj je polazište u molekularnoj strukturi kao temelju čuvanja i prijenosa informacija (npr. struktura DNA). Priložena shema (slika 16) jasno pokazuje kako informacijski koncept molekule oblikuje i povezuje područja u kojima se odvijaju dramatična otkrića i primjene. Definitivno se može reći da je “*data mining*” daleko nadmašio svoju prvotnu namjenu i proširio naše spoznaje i mogućnosti.



Slika 16 – Znanost informacijskih i molekularnih sustava

Fig. 16 – Science of informational and molecular systems

## Literatura References

1. H. M. Berman, K. Heinrich, H. Nakamura, *Nature Structural Biology* **10** (2003) 980.
2. H. M. Berman, O. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, B. Schneider, *Biophys. J.* **63** (1992) 751.
3. A. Belsky, M. Hellenbrandt, V. L. Karen, P. Luksch, *Acta Cryst.* **B58** (2002) 364.
4. P. S. White, J. R. Rogers, Y. Le Page, *Acta Cryst.* **B58** (2002) 343.
5. a) F. H. Allen, *Acta Cryst.* **B58** (2002) 380; b) F. H. Allen, S. Motherwell, *ibidem*, 407.
6. F. H. Allen, R. Taylor, *Chem. Soc. Rev.* **33** (2004) 463.
7. F. H. Allen, R. Taylor, *Chem. Commun.* (2005) 5135.
8. S. R. Hall, F. H. Allen, I. D. Brown, *Acta Cryst.* **A47** (1991) 655.
9. I. D. Brown, B. McMahon, *Acta Cryst.* **B58** (2002) 317.
10. A. Višnjevac, B. Kojić-Prodić, *Kem. Ind.* **51** (2002) 425.
11. M. Čurić, D. Babić, A. Višnjevac, K. Molčanov, *Inorg. Chem.* **44** (2005) 5975.
12. J. D. H. Donnay, D. Harker, *Am. Mineral.* **22** (1937) 446.
13. I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson, R. Taylor, *Acta Cryst.* **B58** (2002) 389.
14. D. G. Watson, *J. Res. Natl. Inst. Stand. Technol.* **101** (1996) 227.
15. Z. Štefanić, Doktorska disertacija, Prirodoslovno-matematički fakultet Sveučilišta u Zagrebu, 2004.
16. K. Molčanov, B. Kojić-Prodić, M. Zehl, G. Allmaier, Fourteenth Croatian-Slovenian Crystallographic Meeting, Book of Abstracts (2005) HKD/HKZ, Zagreb.
17. I. J. Bruno, J. C. Cole, J. P. M. Lommerse, R. S. Rowland, R. Taylor, M. L. Verdonk, *J. Comp.-Aided Mol. Des.* **11** (1997) 525.
18. M. L. Verdonk, J. C. Cole, R. Taylor, *J. Mol. Biol.* **289** (1999) 1093.
19. I. C. Hayes, A. J. Stone, *Mol. Phys.* **53** (1984) 83.
20. G. Jones, P. Willet, R. C. Glen, *J. Mol. Biol.* **245** (1995) 43.
21. G. Jones, P. Willet, R. C. Glen, A. R. Leach, R. Taylor, *J. Mol. Biol.* **267** (1997) 727.
22. M. Hendlich, *Acta Cryst.* **D54** (1998) 1178.
23. J. P. M. Lommerse, S. L. Price, R. Taylor, *J. Comput. Chem.* **18** (1997) 757.
24. K. Molčanov, B. Kojić-Prodić, N. Raos, *Acta Cryst.* **B60** (2004) 424.
25. G. A. Jeffrey, *An Introduction to Hydrogen Bonding*, Oxford University Press, Oxford, 1997.
26. T. Steiner, *Angew. Chem. Int. Ed.* **41** (2002) 48.
27. I. Vujasinović, K. Molčanov, Z. Štefanić, J. Veljković, K. Mlinarić-Majerski, B. Kojić-Prodić, XIX. Meeting of Croatian Chemists and Chemical Engineers, Book of Abstracts (2005) HDKI, Zagreb.
28. I. Vujasinović, J. Veljković, K. Mlinarić-Majerski, K. Molčanov, B. Kojić-Prodić, *Tetrahedron* **62** (2005) 2868.
29. J.-M. Lehn, *PNAS*, **99** (2002) 4763.

## SUMMARY

### 40<sup>th</sup> Anniversary of The Cambridge Crystallographic Data Centre Dedicated to Deposition of Data Related to Crystal and Molecular Structures, "Cambridge Structural Database"

B. Kojić-Prodić and K. Molčanov

The article is dedicated to 40<sup>th</sup> anniversary of The Cambridge Crystallographic Data Centre (CCDC), the world-known centre (<http://www.ccdc.cam.ac.uk>) responsible for deposition and control of crystallographic data, including atomic coordinates that define the three-dimensional structures of organic molecules and metal complexes containing organic ligands. Cambridge Structural Database (CSD), one among the first established electronic databases, nowadays is the most significant crystallographic database in the world. CSD has about 400,000 deposited structures. The use of the extensive database, which is growing rapidly, needs support of efficient and sophisticated software for searching, analysing and visualising structural data. The seminal role of CSD in the research related to crystallography, chemistry, material sciences, solid state physics and chemistry, life sciences, pharmacology, and in particular in drug design, has been documented in more than 1300 scientific papers. The important issues of CCDC are the accuracy of deposited data and development of software that enables a wide variety of applications. Such developing project requires highly competent team of experts; thus the article brings into focus the scientific approach of the team based on the long tradition in crystallography, modelling and informatics. The article is not dedicated to 40<sup>th</sup> anniversary of the centre only, but it also reveals how Cambridge Structural Database can be used in the research and teaching. The use of electronic media and computer graphics makes "data mining" very efficient and useful but also esthetically appealing due to the molecular architecture.

At the Rudjer Bošković Institute, Zagreb, Croatia there is The National Affiliated Centre of Cambridge Crystallographic Data Centre responsible for communication and dissemination of CSD in Croatia, Slovenia and Macedonia.

The use of CSD is illustrated by two examples performed and published by the presenting authors: a) the analysis of the less-common hydrogen bonds with the ester oxygen atom as a proton acceptor, and b) topological analysis of tubular assemblies of macrocyclic polythianes extensively described in the references 24 and 28.